Toward Annotating Commonsense Inferences in Text

Guide for annotators in the TACIT project

**Ernest Davis, Dept. of Computer Science, New York University**

# Overview

The objective of the TACIT project is to identify and characterize, as far as is possible, *all* of the commonsense knowledge and all of the commonsense reasoning that would be needed to understand a small collection of short texts. The primary goal of the project is to explore and map out the space of commonsense inferences that arise in a natural task. A secondary goal is to study the ways in which

Specifically:

* We have chosen a collection of short (3 to 6 sentence) texts of different kinds.
* We are examining these texts carefully to see what inferences implicit in the text need to be made in order to understand the text, and what background world knowledge would be required to make these inferences
* We are characterizing the inferences by categorizing them along three primary dimensions: linguistic significance, category of inference, and domain. There are also a number of secondary dimensions.
* The terms of this characterization ― the dimensions and the categories within each dimension ― are original to this project. Thus, another major part of this project is to develop and refine the terms of this analysis: to make these categories as well-defined, comprehensive, and useful as possible.

Obviously, in our current state of understanding of commonsense reasoning and its relation to natural language understanding, all aspects of this analysis are poorly defined and nebulous. We (that is, the state of the art) have no systematic way of identifying a commonsense inference when it is encountered in text. We have no well-defined way to distinguish an inference that is critical for the understanding of a text from one that is peripheral or of individuating separate inferences. Some of the categories that we are using are drawn as needed from the AI and linguistics literature, but many are our own inventions, and we have no criteria on which to base the legitimacy of a category, other than that we have some examples of it. In short the project is not tied down at any corner; what the inferences are, what the categories are, and how a given inference should be assigned to a given category, are all largely up for grabs. The project rests on the hope that, by working through the process of identifying inferences and categorizing them, all of these issues will gradually become clearer, and a reasonably well-defined theory will gradually take form and emerge from the mists. So far, this is just a hope.

For this reason, it is unrealistic to expect any substantial degree of inter-annotator agreement, except in very clear-cut cases, such as lexical ambiguity or coreference resolution. It would be difficult even to find a good measure of interannotator agreement in identifying the set of inferences to be made. Given a fixed set of inferences, it would certainly be possible to use standard measures of interannotator agreement in assigning categories; we may do this at some later date.

The annotator facing a text begins by putting herself into the mind-set of a hypothetical computer program which has a thorough knowledge of English vocabulary, syntax (grammatical rules), and semantics (the meaning expressed by grammatical relations), but no knowledge of anything that is true in the outside world that is not directly expressed in the text, nor of the discourse conventions that govern how writers and speakers structure a text --- what they choose to include or to leave out. The question to be answered is, what in the meaning of the text would such a program be confused about, misunderstand, or simply miss altogether? Having found such a gap, the annotator must then try to answer the following questions:

1. What *background knowledge* does the human reader use in order to resolve the gap? We are primarily interested here in knowledge of facts about the external world and discourse conventions, not in facts about the language (e.g. the meanings of idioms).
2. How would you categorize the *domain* of the background knowledge in (1)?
3. Why is it *important* for the human reader to fill this gap or make this connection? We call this the “linguistic significance” of the inference. As we will discuss below, the more specific this can be made, the better.
4. What is the logical form of the fact being inferred?

# Examples

We illustrate with a pair of examples of inferences and their characterization from our corpus (we have chosen comparatively clear-cut examples.)

In the first news story text (see section 3 below) about the theft of the Mona Lisa in 1911, the first two sentences read as follows:

On a mundane morning in late summer in Paris, the impossible happened. The Mona Lisa vanished.

The first, third, and fourth inferences associated with these two sentences read as follows:

**Inference 1 :** In "the impossible happened", "impossible" is hyperbolic, not literal. What is meant is "a very improbable event".

**Specific text being explicated:** "the impossible happened"

**Background:** An impossible event cannot happen.

**Category of Inference:** ( PropertyOf = Unlikely ; Event = "the impossible" ; )

**Domain:** Theory of necessity and possibility.

**Linguistic Significance:** Interpret non-literal text.

**Question:** How likely did the event under discussion seem before it occurred? **Right answer:** Quite unlikely. **Wrong answer:** Impossible. **Wrong answer:** Likely. **Wrong answer:** Certain.

**Question:** How likely is it now that the event under discussion occurred? **Right answer:** Certain. **Wrong answer:** Likely. **Wrong answer:** Quite unlikely. **Wrong answer:** Impossible.

**\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\***

**Inference 3 :** In "The Mona Lisa vanished", "vanished" is metaphorical, not literal. What is meant is "The Mona Lisa became absent from its proper place".

**Specific text being explicated:** "The Mona Lisa vanished"

**Background:** Physical objects rarely literally vanish.

**Category of Inference:** ( Existence ; Event = Mona Lisa became absent ; )

**Domain:** Spatial and physical knowledge

**Linguistic Significance:** Interpret non-literal text.

**Question:** What actually happened to the Mona Lisa? **Right answer:** The Mona Lisa unexpectedly became missing from its usual place. **Wrong answer:** The Mona Lisa became invisible.

**\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\***

**Inference 4 :** The event of the Mona Lisa leaving its place and the event judged to be impossible in sentence 1 are the same event.

**Specific text being explicated:** "... the impossible happened. The Mona Lisa vanished"

**Background:**

1. It is important that valuable objects remain where they are supposed to be, and great efforts are made to ensure that they do so. Therefore, it is considered highly improbable that a valuable object will leave its place, other than under the supervision of the authorities responsible for it.
2. A painting in a museum is a valuable object.
3. Paintings in a museum are under the supervision of the museum administrators

**Compare:** "… the impossible happened. A bar of soap had vanished from the men's bathroom at the Louvre."

**Category of Inference:** ( Identical ; Event = "the impossible" ; Event = "Mona Lisa vanished" ; )

**Domain:** Organizations. Property.

**Linguistic Significance:** Coreference resolution.

**Additional Linguistic Clues:** The metaphorical "vanished" fits with the hyperbolic "impossible"; it would be literally impossible for the Mona Lisa to literally vanish.

**Question:** What is the connection between “the impossible happened” and “The Mona Lisa vanished”? **Right answer:** The Mona Lisa vanishing is the impossible event that happened. **Wrong answer:** First the impossible happened, then the Mona Lisa vanished.

**Question:** Why was the event under discussion considered very unlikely? **Right answer:** Because museums try hard to make sure that their valuable artworks are always in the proper place. **Wrong answer:** Because paintings do not usually vanish.

**\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\***

Hopefully these are reasonably self-explanatory. The first two inferences that need to be made are that “impossible” and “vanished” are figurative, not literal. The third inference is that the phrases “the impossible happened” and “the Mona Lisa vanished” refer to the same event. The linguistic significance of the first two is to interpret non-literal text; the linguistic significance in the third is coreference resolution (determining that two entities mentioned in the text are the same). The first requires general knowledge that impossible things cannot in fact happen; the domain of this fact comes is the general theory of necessity and possibility. The second requires the more specific knowledge that physical objects rarely literally vanish; this comes from a physical theory. We categorize the conclusion in the first as the inference that the event (whatever it is) denoted as “the impossible” has the property of being unlikely. We categorize the conclusion in the second as an inference that the event of Mona Lisa becoming absent occurred (existed). The format we use for these ― e.g. “( PropertyOf = Unlikely ; Event = "the impossible" ; )” ― is explained in the next section.

The third inference is substantially more complex. Having interpreted “the impossible happened” as “a very unlikely event occurred” and having interpreted “the Mona Lisa vanished” as “the Mona Lisa became absent from its usual place in an unexpected way”, the reader must now connect the two. This involves understanding why it is that the unexpected absence of the Mona Lisa would be considered so extremely unlikely; as the sentence introduced as a point of comparison illustrates, if a bar of soap unexpectedly became absent, one would hardly describe that at “the impossible happening” except as a joke. This understanding thus depends on an understanding of the value of famous paintings and the care that is taken to make sure that their whereabouts are always known to the responsible authorities. We have formulated this knowledge in the background facts numbers 1-3; obviously, the individuation as separate facts is somewhat arbitrary. We characterize the home domain of these facts as partly in the theory of property and partly in the theory of organizations.

The linguistic significance of this inference comes under the category of coreference resolution; we need to determine that two separate (and quite different) phrase in the text in fact refer to the same entity (the Mona Lisa becoming absent). We categorize the type of inference as the identification of two different events (“identification” in the sense of “showing that the events are identical”, not in the sense of “determining the identity”).

We note further, under “Additional linguistic clues” that this interpretation receives further support from the fact that the writer is continuing the same figure of speech; having said that the event is impossible, he describes it in terms that are, in fact, impossible.

With each inference, we include one or more multiple-choice questions to test whether the inference has been adequately carried out.

The value of all this analysis is to make a small incremental contribution to our understanding of commonsense reasoning and its role in text interpretation. We have shown that achieving basic competence in understanding these two naturally occurring sentences requires knowing the five commonsense facts, or something equivalent, from the domains of necessity and possibility, physical knowledge, knowledge of property, and knowledge of institutions, and involving reasoning of these two different categories. If we do enough of this kind of analysis, we should begin to get some sense of general space of commonsense knowledge and commonsense reasoning needed for the interpretation of various kinds of text.

# Current State of the Corpus

The home page for the project is located at <http://www.cs.nyu.edu/faculty/davise/annotate/Tacit.html>

As of September 1, 2014, six texts have been analyzed all from news stories. A total of 139 inferences have been identified and characterized.

# Characterization of Commonsense Inferences

## Dimensions

We characterize inferences along three dimensions.

*Linguistic Significance* is the role that the inference plays at the level of text processing (e.g. lexical or semantic disambiguation) or text processing (e.g. explicate causal structure). In cases where an inference may have multiple significances of this kind, we generally prefer the lowest level; e.g. prefer "lexical disambiguation" over "explicate causal structure".

*Domain* is the knowledge domain of the background knowledge.

*Category of Inference* categorizes logical form of the conclusion. Here we use a semi-formal structure. The inference is categorized in terms of an operator, which is a relation, and arguments, which are entities. For each operator and argument we specify:

* A general category, from a fairly fixed list. For instance *PropertyOf* and *Existence* are categories of relations; *Person* and *Event* are categories of entities.
* A specific value of these categories, with the exception of certain basic relations. For instance in inference 1 above *PropertyOf* has the value *Unlikely* and *Event* has the value *“the impossible”* (in quotes, to emphasize that this is a reference to the text). In inference 3, *Existence* has no specific value, and *Event* has the value *Mona Lisa became absent*. These values have no particular structure; they are written in abbreviated English which hopefully is intelligible to the human reader.
* Both operators and arguments may have the modifier *Not.* Arguments may additionally have the modifier *Multiple.* For example, inference 5 for this story is the inference that the Mona Lisa was not removed by the museum administration. The operator for this is *Not RoleIn = Actor*; the arguments are *Multiple Person = Administration* and *Event = Remove Mona Lisa.*

In our descriptions below of *Linguistic Significance* and *Domain*, we give the name (in bold face), a brief description, an example, and an enumeration of the inferences in the category. In our description of *Category of Inference*, we enumerate the categories of relations and of entities, with descriptions, and an example; and an enumeration of the inferences for each category of relation. When the name of the category is self-explanatory, the description just repeats it. When there are two or more categories that are clearly close to one another, we add an explanation of the intended distinction between them. We include here only categories of inferences that we have encountered in our texts; the absence of a category from this list means only that we have not yet run across it.

Abbreviations: Nk.n = News Text k, inference n. The cross-references here are outdated,

## Linguistic Significance

**Abstract frame.** *Description:* A “frame” (Minsky, 1975) is a stereotypical structure consisting of standard components related in standard ways. For example, the “house” frame contains rooms, doors, windows, and so on, arranged in the usual way. The inference here involves identifying a frame from a component.  *Example:* In N3.6 we infer that the event "workers walked off the job" is an instance of the frame "strike". Inferences: N3.6

**Characteristic of an entity.** *Description:* Infer that an entity mentioned in the text has a particular property. *Example:* In N2.2, infer that "Angels In Waiting" is a non-profit organization. (Owing to a somewhat confusing syntax, this is not clearly stated in the text.) *Inferences* N2.2

**Clarify misleading syntax.** *Description:*In cases where the syntax is misleading --- e.g. where a modifier is syntactically attached to an constituent other than the one actually modified --- correct the interpretation. *Example:* In N2.2, "nonprofit" actually characterizes the organization "Angels in Waiting" rather than the particular group of nurses that Suleman fired (however, see discussion there.) Inferences: N2.2

**Clarify vague expression.** *Description:* Determine the actual significance of a vague expression. *Example:* In N3.1, "nearly a million" presumably means more or less "between 800,000 and 1,000,000". Inferences:N3.1

**Coreference resolution.** *Description:* Determine what previously mentioned item a noun phrase (often a pronoun) refers to. *Example:* In N3.17 the word "They" in "They complied with this order" refers to the workers. Inferences:N1.4, N1.6, N1.12, N2.13, N3.8, N3.17

**Counter argument**. *Description: Example: Inferences*: N6.10

**Ellipsis.** In a text with an ellipsis, restore the missing words.

**Explicate causal structure.** *Description:* This is by far the most common category in this dimension in the examples we have considered. The inference is made in order to explain the causal structure of the events described in the text. It would probably be desirable to refine this into subcategories of different kinds of causal connection; however, we do not currently have a good theory to support that. *Example:* In B1.11 and 12, to understand the sentence "[A] river may change course and divide a population of animals that cannot cross it" the reader must infer that initially the region inhabited by the species lies on one side of the river; then the river changes to a new course that divides that region into two, leaving different subpopulations on opposite sides of the river. *Inferences:*B1.2, B1.3, B1.4, B1.5, B1.9, B1.10, B1.11, B1.12, B1.13, B2.4, B2.5, B2.8, B2.9, B3.1, N1.5, N1.11, N1.14, N1.16, N1.18, N1.19, N1.20, N1.22, N2.8, N2.12, N3.4, N3.9, N3.14, N3.15, N3.19, N3.21

**Find case filler.** *Description:* Infer the filler for a case of some word that appears in the text. *Example:* ADD EXAMPLE*. Inferences:* N1.2, N1.23, N1.24, N2.6, N2.7, N2.15, N2.19, N3.10, N3.12, N3.20

**Identify entity.** *Description:* Identify the referent of a noun phrase with an entity that has not been previously mentioned in the text. *Example:* ADD EXAMPLE. *Inferences:* N1.17

**Interpret non-literal text.** *Description:* Explain the meaning of a figure of speech or other non-literal text. *Distinction:* As is well known, the borderline between a figure of speech and an alternative meaning is hazy; it depends on how fixed the usage is. (There are some linguistic tests, but these do not always give the same answer.) *Example:* In N1.1 "the impossible happened" is a hyperbole for "an entirely unexpected event happened." *Inferences:*N1.1, N1.3, N1.26, N2.6, N3.3

**Lexical disambiguation.** *Description:* Determine the meaning of an ambiguous word. *Example:* In N2.3 determine that "charged" means "assigned a duty". *Inferences:* N1.17, N2.3, N2.18, N3.5, N3.7, N3.11

**Motivation analysis.** *Description:* Determine the motivation of an action, or of some aspect of an action. *Example:* In N1.15 determine that the thief stole the Mona Lisa when the museum was closed rather than when it was open in order to escape detection. *Inferences:*N1.15, N1.21, N1.25, N1.27, N2.9, N2.10, N2.16, N2.21, N3.18

**Named entity interpretation.** *Description:* Characterize a named entity. *Example:* In N1.8, infer that "Mystical Marriage" and "Allegory of Alfonso d'Avalos" are works of art. *Inferences:*N1.7, N1.8, N2.5

**Noun-phrase semantic structure.** *Description:* Determine the semantic structure of a complex noun-phrase; that is, the actual relation between its components. In compound nouns (phrases that are just a sequence of nouns) in particular, the syntactic structure can be highly ambiguous and can indicate little or nothing about the actual relation between the words. *Example:* In N2.5 determine that "CNN affiliate KTLA" means "KTLA, which is affiliated with network CNN". *Inferences:*N2.5

**Qualification on an event.** *Description:* Infer a qualification on a phrase denoting an event that significantly limits or changes its meaning. *Distinction:* The category is close to "Infer characteristic of an event". The intended difference is analogous to the distinction between a restrictive and a non-restrictive clause. A "characteristic" of an event is just an important feature, whereas a "qualification" is intrinsic to the characterization of the event. *Example:* ADD EXAMPLE *Inferences:* N1.31.

**Qualification on a modifier.** *Description:* Infer a qualification on a modifier in the text that significantly limits or changes its meaning. *Distinction:* This category is close to both "Find case filler" and to "Semantic disambiguation". The intent is that it is more intrinsic to the meaning of the word ("isolated") than "Find case filler" but supplies additional information, unlike "semantic disambiguation". *Example:* *Inferences:*

**Reference separation**. *Description: Example: Inferences*: N4.11

**Relate example to general description.** *Description:* When a general description is followed by a concrete example, infer how the example in facts illustrates the general description (e.g. by matching the components of the example to the components in the description). *Example:* *Inferences:*

**Relation between events.** *Description:* Infer a relation between two events. *Example:* In N2.20 infer that successive events of discharging the babies occur at at intervals on the order of weeks. *Inferences:*N2.20

**Sentiment analysis.** *Description:* Infer the attitude of either the author of the text or of a person mentioned in the text toward some event or entity. *Example:* In N3.13, infer that when the workers said they had been locked out, they were seeking to assign the responsibility for the work cessation to the management. *Inferences:*N2.11, N3.13

**Semantic disambiguation.** *Description:* Disambiguate a semantically ambiguous phrase; that is, a phrase where the grammatical structure is clear, but the actual relations between the elements is unclear. *Example:* In B3.3 the phrase "the organs that exchange gasses, absorb nutrients, and dispose of wastes" refers to three different kinds of organs rather than one kind of organ that carries out all three functions. *Inferences:*B1.8, B2.7, B3.3, N1.2, N1.9, N2.17, N3.16

**Source of information.** *Description:* Infer the source of a fact reported in a narrative. This can be important in evaluating reliability, determining meaning, or constructing a causal structure. *Example:* In N3.2 infer that the estimate of the number of people affected derives from the average volume on the transit system. *Inferences:*N3.2

**Syntactic disambiguation.** *Description:* Disambiguate a syntactically ambiguous phrase. *Example:* ADD EXAMPLE. *Inferences:* N1.10, N1.28, N1.29, N1.30, N2.4

**Temporal sequence**. *Description: Example: Inferences*: N4.11

## Domain

**Agents and actions.** *Description:* Use general knowledge about how agents choose actions and how they carry them out. *Example:* In N1.21, use the facts that it is harder to carry a larger item than a smaller one and that people prefer to carry out an easier action. *Inferences:*N1.14, N1.15, N1.18, N1.21, N1.22, N3.3

**Art.** *Description:* General knowledge about works of art. *Example:* In N1.8 use the fact that things with proper names in museums, other than people, are generally works of art. *Inferences:*N1.6, N1.8, N1.9

**Biology**. *Description:* Use knowledge of biology. *Example*: In N4.3, use the fact that hairs are part of people. *Inferences:* N4.3

**Communication.** *Description:* Use knowledge about communication and how it relates to the mental states of speaker and hearer. *Example:* In N2.10 use the fact that a negative report about a person generally requires some evidence. *Inferences:*N1.22, N1.26, N1.31, N2.10, N2.6, N2.12, N2.21, N3.2, N3.3

**Default goals.** *Description:* Use knowledge of the basic goals that are common to people. *Example:* In N2.16, use the fact that people generally prefer to pay less rather than more for a given service. *Inferences:*N2.9, N2.16

**Default properites of categories.** *Description:* Use knowledge about default properties of particular categories. *Example:* In N1.17 use the fact that paintings are often in frames. *Inferences:* N1.10, N1.17, N1.21

**Default spatial relations.** *Description:* Use knowledge about characteristic spatial relations. *Example:* In N1.16 use the fact that radiators are generally in buildings. *Inferences:*N1.16

**Discourse convention.** *Description:* Use a fact about how texts are organized (i.e. facts about texts rather than object-level facts about the world). *Distinction:* Gricean inference is a distinguished subcategory. *Example:* In N1.10 use the fact that the location of objects is usually specified in terms of its position relative to other objects roughly the same size and similar category. *Inferences:*N1.10, N1.24, N3.1

**Folk psychology (other).** *Description:* Use folk psychological knowledge (i.e. general, non-scientific knowledge about how people think and how that affects their behavior) of a kind that doesn't come into the other categories here. *Example:* In N2.19, use the fact that it may be difficult for a parent to adjust to the sudden arrival of eight new babies. *Inferences:*N2.19, N2.20

**Geography**. *Description:* Use knowledge of geography. *Example*: *Inferences:*

**Gricean inference.** *Description:* Interpret the text using the assumption that the author is following the [Gricean rules](https://en.wikipedia.org/wiki/Cooperative_principle#Grice.27s_Maxims). The most important of these here are the maxims of quantity: “Make your contribution as informative as is required (for the current purposes of the exchange)” and “Do not make your contribution more informative than is required.” *Example:* In N2.17, from the fact that the text says that four of the octuplets are at home, infer that the other four are not at home. *Inferences:*N2.17

**Knowledge and belief.** *Description:* Use knowledge about the knowledge and belief of agents. *Example:* Use the fact that, if a person carries out an action themselves with entirely predictable effects, they generally do not find those effects surprising. (There are of course many exceptions, but this is the default.) *Inferences:*N1.5, N1.26, N1.27, N1.31, N2.10, N2.12

**Importance of narrative information** *Description:* Interpret the text as conveying important rather than unimportant information (this is also a Gricean inference). *Example:* In N1.2 use the fact that it is generally important to know the location of an event; and generally unimportant to know the time of an event at some other location to determine the meaning of "morning ... in Paris". *Inferences:*N1.2

**Naming conventions.** *Description:* Use conventions for assigning proper names to entities. *Example:* In N2.5, use the fact that television stations have a 4 letter identifier starting with K. *Inferences:*N2.5

**Narrative focus or point of view** *Description:* Use the focus or point of view of the narrative. *Example:* In N3.8, the referent of "their busses" is resolved by using the fact that the commuters rather than the workers are the focus of the previous sentence. *Inferences:*N2.5

**Necessity and possibility.** *Description:* Use a basic theory of what it means for something to be possible, impossible, or inevitable. *Example:* In N1.1 use the fact that impossible events do not happen, by definition. *Inferences:*N1.1

**Organizations.** *Description:* Knowledge about organizations, how they work, and how they interact with people. *Example:* In N1.7 and N1.8, use the facts that mistreatment of children can be reported to child-welfare authorities, and that the result can be that the authorities remove the children. *Inferences:*N1.4, N1.5, N1.7, N1.12, N1.23, N1.24, N1.25, N1.26, N1.27, N1.28, N1.29, N2.2, N2.5, N2.6, N2.7, N2.8, N2.10, N2.13, N2.18, N2.19, N2.21, N3.9, N3.10, N3.11, N3.14, N3.15, N3.18

**Property.** *Description:* Use knowledge about property relations. *Example:* In N1.11, use the fact that when a valuable object becomes missing, it is likely to have been stolen. *Inferences:*N1.4, N1.11

**Readers' knowledge.** *Description:* Use default assumptions about typical reader's knowledge, attributed to the writer. *Example:* In N1.30, use the fact that typical readers do not know the names of museum directors. *Inferences:*N1.30

**Social relations.** *Description:* Knowledge of characteristic relations between people. *Example:* In N1.15 use the fact that, if one person sees another engaging in an illegal activity, they may stop them, or alert some one else. *Inferences:*N1.12, N1.13, N1.15, N1.18, N2.1, N2.3, N2.4, N2.11, N2.12, N2.15, N2.16, N3.2, N3.4, N3.5, N3.6, N3.12, N3.13, N3.16, N3.17, N3.18, N3.20, N3.21

**Spatial and physical knowledge.** *Description:* Use a combination of knowledge about spatial and physical relations. *Example:* ADD EXAMPLE. *Inferences:* N1.3, N1.10, N1.14, N1.18, N1.19, N1.20, N1.21, N3.3

**Temporal reasoning.** *Description:* Use basic knowledge of temporal relations and how they combine. *Example:* In N1.12, N1.13, N1.14, combine the various temporal constraints to obtain a presumed ordering of the actions that the thief took. *Inferences:*B2.4, N1.12, N1.13, N1.14, N1.18, N1.20

## Categories of Inference

#### Arguments:

**Aspect:** *Description:* Some particular aspect of an entity. This is a rare category with only one example, thus far. *Example:* In N1.17, the inference requires explaining the motivation of the thief to steal the painting on Sunday or Monday rather than some other day. The representation is “( Motivation ; Aspect = On Sunday or Monday ; Event = Remove Mona Lisa from Louvre ; Not Event = Catch Thief ; Person = Thief ; )” That is, the motivation for the thief of the aspect ``On Sunday or Monday’’ of the event “Remove Mona Lisa from Louvre” is that the event “Catch Thief” should not occur.

**Event:** *Description:* An event or action, other than a speech act. Events may be characterized at any level of abstraction and with any degree of completeness of specifications of the roles involved.  *Example:*In N.17, the event is “Remove Mona Lisa from Louvre.”  *:*

**Object:** *Description:*An object or place, other than a person. *Example:*The Mona Lisa. The Louvre. Paris.

**OtherEntity:** *Description:*Any entity that does not fall into one of the other categories*. Example:* There is thus far one example: In N5.7, the “challenge” posed to the United States by the fact that the most effective rebel force is an Islamic extremist one.

**Person:** *Description:* People.  *Example* In News1 the thief who stole the Mona Lisa; in News2 Suleman.. Note that organizations, institutions etc. such as the museum administration and the news agencies in News1 are categorized as “Multiple Person”.

**Proposition:** *Description:* A proposition that is not a simple state or event. Used as an argument to “Motivation”, “ContentOf”, and “Believe”. *Example:*Thus far, there is only one example: In N5.8, the proposition that the most effective rebel group is an Islamic extremist one.

**SpeechAct:** *Description:*A speech act or communicative act. *Example:* In News3, the rulings by the mayor and labor board.

**State:** *Description:* A state that holds at a given time. The distinction between State and Event is the vague distinction between something being true at a time and something happening. *Example:* In News4, the romantic relation between Gibson and Grigorieva. In News5 the facts that the Nusra front is small and the fact that it is effective are each states.

#### Relations:

Except where noted, these relations do not have specific values associated; the annotation just uses the name of the category

**Authorized:** *Description:*A person is authorized to carry out an action. *Example:*In N3.13, the mayor and labor relations board are authorized to issue rulings on the legality of strikes. *Inferences:* N3.13.

**Believe:** *Description:*A person believes an event, state, or proposition. *Example:*In News2, Suleman believes that the nurses are spying on her. *Inferences:*

**CausalRelation:** *Description:*There is a causal relation between two or more entities, which may be states, events, or propositions. *Example:* In N5.12, the state that the Nusra Front is bold and skillful causes the state that it is effective. *Inferences:*

**ContentOf:** *Description:*The content of speech act is an event, state, or proposition. *Example:* In N2.7, the content of the nurses’ hypothetical report to the authorities is that Suleman is an inadequate parent. *Inferences:*

**Emotion:** *Description:* A person has a specified emotion. *The value is specified.* *Example:*In N4.5, Gibson is under stress. In N4,6, he is not under extreme stress. *Inferences:*

**Ethics:** *Description:* Characterize the ethical standing of an event or state. *The value is specified.*  *Example:* In N2.11, the event of the nurses spying on Suleman would be unethical. *Inferences:*

**Existence:** *Description:*A person or object exists, or an event occurs. “Existence” here is intended in a weak sense; hypothetical events may “exist” if they are under consideration. This category does not have specific values.  *Example:* News1 involves inferring the existence of a number of events: The event of the Mona Lisa going missing, the event of it being stolen, the event of it being removed from its frame, the event of the thief putting the frame behind the radiator. *Inferences:*

**Goal:** *Description:*A person has a goal, which is a state or event. *Example:* In N2.9, Suleman has the goal that the authorities should not intervene. (The representation is “(Goal; Person=Suleman; Not Event=Authorities intervene)”.) *Inferences:*

**Identical:** *Description:* Two entities are identical. *Example:* In News4, Mel Gibson, the Autralian actor, and the Oscar-winning Braveheart star are identical. In News1 the event described in the first sentence as “the impossible” and the event described in the second as “the Mona Lisa vanished” are identical. *Inferences:*

**Identify:** *Description:* Identify an entity that the text only partially characterizes. *Example:* In N1.19 identify “the empty frame” as the frame of the Mona Lisa. In N2.19 identify “the adjustment” as Suleman’s adjustment to dealing with a family of 14 children. *Inferences:*

**Motivation:** *Description:* A motivation for Entity1 (usually an event) is motivation Entity2 relative to Person. *Example:* In N1.24, the motivation for “Take Mona Lisa out of frame” is “Take Mona Lisa out of the Louvre” relative to the thief. *Inferences:*

**OtherRelation:** *Description:* Any relation not otherwise listed here.  *The value is specified. Example:* In N1.9, the relation “Created” holds between Corregio and “Mystical Marriage”. In N2.18, the relation “Disjoint” holds between the octuplets and the six other children. *Inferences:*

**PartOf:** *Description:* Entity1 is part of Entity2. Usually either both entities are objects or people or both are events. *Example:*In N4.4, the gray hairs are part of Mel Gibson. N4.14, the event “Walk on the red carpet” is part of the event “Attend the movie premiere.”  *Inferences:*

**Perceive:** *Description:* A person perceives another person, object, state, or event. *Example:* Suleman (claimed that) she perceived the event of the nurses spying on her. *Inferences:*

**PropertyOf:** *Description:* An entity, of almost any category, has a specified property. *The value is specified.* *Example:*In N1.8, “Mystical Marriage” has the property “WorkOfArt”. *Inferences:*

**QualificationOnProperty:** *Description:* A stated property holds with respect to some reference class or in some particular context. Because of the logical complexity involved here, we simply state this with no arguments, as a category. *Example:* In N1.27 the text states that Le Temps was the leading morning newspaper. What is meant in that it was the leading morning newspaper in Paris. In N3.9, the text states that the service that continued was of limited use. What is meant is that it was of little use to the commuters. *Inferences:*

**RoleIn:** *Description:* One entity (of almost any kind) plays a specified role in some other entity (usually an event). *The value is specified.* *Example:*In N2.4, the nurses have the role “Assist” in the event “Care for children”. In N4.10, Gibson has the role “Participant” in the event “Divorce” *Inferences:*

**SocialSignificance:** *Description:* An entity (state or event) signifies some social state, event, or proposition. *Example:* In N4.2, the event of Gibson and Grigorieva holding hands signifies the state that they have a romantic relationship. *Inferences:*

**SpatialRelation:** *Description:* Some number of objects or people lie in some spatial relation.*The value is specified.* *Example:* In N1.12 the Mona Lisa ordinarily hangs between “Mystical Marriage” and “Allegory of Alfonso d’Avalos”.  *Inferences:*

**TemporalRelation:** *Description:*One or more events or states satisfy a temporal relation. *The value is specified.* *Example:* In N1.22 the event “Take the Mona Lisa out of its frame” precedes the event “Place the frame behind the radiator.” *Inferences:*

# Question and Answer

With each inference, we present one or more multiple-choice question and answer. The right answer and wrong answers are indicated in the corpus.

The question is intended to be based on having read the passage up through the “Specific text” associated with the inference, and is worded correspondingly. In some cases, the text further on in the passage gives further information about the answer to the question, and in some cases, later questions implicitly give the answers to earlier ones. Thus, the questions should be viewed as being given sequentially, after the corresponding part of the text has been read.

In formatting the wrong answers, we have been careful to choose them and word them so that they are unequivocally wrong; they do not merely fail to capture the meaning of the text, they are manifestly untrue. In some cases, this limits the precision with which we can test the understanding of the text. Also, some of the wrong answers can be excluded using other kinds of knowledge than we have enumerated under “Background”.

It is certainly possible that there are stylistic or other clues in the wording of the question and answer that would allow a program to select the right answer without actually understanding the text. It would be extremely difficult to prevent this, and we have not made any effort to mitigate this. This corpus therefore should not be used in applications where “cheating” of this kind is a consideration; challenges of other formats such as [the Winograd Schema Challenge](http://www.cs.nyu.edu/faculty/davise/papers/WS.html) are designed to avoid this kind of issue.

# XML format

The annotations are in XML files, one text per file. Unlike most text tagging projects, the annotations are not in the form of tags inserted in the text, as the content of the annotations is much too large to support that. Rather, there is an initial segment consisting of fields that apply to the text as a whole; followed by a sequence of individual annotation.

The XML tags used are listed below (as always in XML, opening tags must be balanced by the corresponding close tag).

<annotatedText> -- begins and ends the file as a whole>

## Introductory segment

The following tags apply to the entire text. They are at the beginning of the file. There is at most one instance of each tag per text.

<corpusNumber> — Identifier of text (e.g. “News Text 1”)

<annotators> — Names of annotators

<source> — External source of text

<text> — Content of text

<generalComment> — General comment on text and interpretation issues it raises (Optional)

<commonKnowledge> — Background knowledge assumed, either of individuals (named entities) or technical knowledge. This should be knowledge that the average reader can be expected to have in advance. (Optional)

<paraphrase> — When useful, a paraphrase of the content of the text (e.g. in B.3). (Optional).

<additional> — Additional comments about the text, to be displayed at the end of the file.

## Individual inferences

The introductory segment is followed by a sequence of individual inferences (annotations).

<annotation> — Overall delimiter for inference.

<infNumber> — Counter for inference (this should, of course, be automated).

<inference> — Description of the inference being carried out.

<specificText > — Specific phrase in the text that requires the inference for its understanding.

<structure> — Logical structure of the inference, in the form of an argument.

<background> — Background fact or facts. If more than one fact, this should have the form <background multiple=”true”>

<fact> — Individual background fact being used. These are within the scope of a <background> tag.

<compare> — Alternative text for comparison. (Optional)

<categoryInf> — Category of inference. This has embedded a relation, delimited with <op>, and one or more entities, delimited with <arg> . <op> may have the modifier not=”true”. <arg> may have either or both of the modifiers not= “true” and multiple = “true”.

<domain> — Domain of background facts.

<significance> — Linguistic significance of inference.

<linguisticClue> — Additional linguistic clues supporting inference (see, for instance, N1.4) . (Optional).

<interaction> — Interaction of this inference with other inferences. This has been largely neglected in the current state of the corpus; that should be fixed. (Optional).

<retrievable> — Include this tag if the fact being inferred could also be gotten directly from online sources. (Optional. This tag has no associated content; that is, you just write <retrievable></retrievable>.)

<remark> — Additional comment (Optional).

<discourseKnowledge> — Used when the inference is made on the basis of pragmatic knowledge about how people express themselves, rather than facts about the external world; e.g. Gricean maxims. (Optional)

<qa> — A multiple-choice question and answer set. This has imbedded a question, delimited with <question>; a correct answer, delimited with <rightanswer>; and one or more incorrect answers, delimited with <wronganswer>.

<question> — The question in a multiple-choice question and answer set.

<rightanswer> — The correct answer in a multiple-choice question and answer set.

<wronganswer> — The incorrect answer in a multiple-choice question and answer set.